

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan dari penelitian Deteksi *Trending Topics* dari Data *Stream Twitter* dengan Pendekatan *Sequential K-Means* adalah sebagai berikut:

1. Model yang telah dirancang dan digunakan pada penelitian terdiri dari empat bagian yaitu sebagai berikut: pertama proses pengumpulan data *twitter*. Kedua, praproses data yang terdiri dari praproses teks, pembobotan kata dan seleksi kata. Ketiga, proses analisis data dengan teknik *clustering*, algoritma yang digunakan yaitu *k-means* dan *sequential k-means*. Keempat, pengolahan informasi yang terdiri dari evaluasi *cluster*, deteksi topik, visualisasi *trending topics* dan evaluasi *trending topics*. Adapun strategi yang telah dirancang dan digunakan pada penelitian ini yaitu batas waktu (*timeout*) pada tahap pengumpulan data ditentukan selama 60 detik, menghapus *hashtags* dari dokumen pada tahap praproses teks karena topik pada penelitian ini adalah kata bebas yang terlepas dari *hashtags*, kata yang dipilih sebagai topik adalah kata benda (*noun*) pada tahap seleksi kata, nilai batasan kemiripan antara data baru dengan *cluster center* ditentukan sebesar 0,1 pada tahap *sequential k-means clustering*, evaluasi *trending topics* dilakukan setiap tiga jam sekali untuk setiap *cluster*, dan topik yang akan dihapus yaitu topik dengan nilai *delta* tidak mencapai 100 *tweets*.
2. Aplikasi deteksi *trending topics* dari data *stream twitter* dengan pendekatan *sequential k-means* dibangun dengan menggunakan bahasa R dan *library* pendukung seperti *ROAuth*, *streamR*, *NLP*, *openNLP*, *openNLPmodels.en*, *tm*, *stringr*, *proxy*, *cluster*, dan *clValid*. Tahapan proses pada aplikasi ini sesuai dengan model yang telah dirancang.
3. Hasil akhir penelitian menunjukkan bahwa sebesar 35.48 persen dari total iterasi memiliki kualitas *cluster* yang baik yaitu dengan nilai *dunn index* lebih dari 0,7.

4. Perbandingan terhadap penelitian yang dilakukan penulis dengan penelitian lain yang serupa telah dilakukan dengan tujuan untuk membandingkan

keunggulan pada setiap penelitian dimana metode yang digunakan untuk deteksi *trending topics* berbeda-beda. Setiap penelitian memiliki keunggulan masing-masing baik itu dari segi kecepatan, keefektifan dan kualitas hasil yang didapat. Metode yang digunakan pada masing-masing penelitian disesuaikan dengan data masukan yaitu ada yang bersifat *offline* dan *online*. Metode untuk deteksi *trending topics* bisa dengan teknik TF-IDF, klasifikasi, *clustering* dan graf jaringan.

5.2 Saran

Berikut ini merupakan saran pada penelitian ini untuk pengembangan lebih lanjut.

1. Penelitian untuk mencari *trending topics twitter* dapat menggunakan algoritma khusus untuk menangani *clustering* data teks *stream* seperti algoritma *Online Clustering of Text Streams* (OCTS) dan *Online Clustering of Text Streams with Merge* (OCTSM) yang ditawarkan pada penelitian (Liu, Cai, Yin, & Fu, 2008).
2. Data yang digunakan untuk penelitian *trending topics twitter* menggunakan *tweet* bahasa Indonesia.
3. Perlu adanya identifikasi *event* dan *non event* seperti pada penelitian (Becker, Naaman, & Gravano, 2011). Sehingga topik yang dihasilkan bukanlah kata yang bersifat umum atau tidak menjelaskan kejadian khusus.